(12) UK Patent Application (19) GB (11) 2 368 936 (13) A

(43) Date of A Publication 15.05.2002

(21) Application No 0103839.7

(22) Date of Filing 16.02.2001

(30) Priority Data
(31) 0027508 (32) 09.11.2000 (33) GB

(71) Applicant(s)
The University of Sheffield
(Incorporated in the United Kingdom)
Western Bank, Sheffield, S10 2JF, United Kingdom

(72) Inventor(s)
James Stone

(74) Agent and/or Address for Service
Harrison Goddard Foote
Fountain Precinct, Leopold Street, SHEFFIELD,
S1 2QD, United Kingdom

(51) INT CL[7]
G06F 17/15

(52) UK CL (Edition T)
G4A ACL

(56) Documents Cited
WO 01/17109 A1 US 5959966 A
US 5825671 A
"Adaptive blind source separation for vurtually any source probability density function" Zarzoso & Nandi Pages 477-488, IEEE Trans on Signal Process, Feb 2000 ISBN 0-7803-5673-X
"Maximum likelihood blind source separation in gaussian noise" Miqeuz & Castedo Pages 343-352 IEEE Signal Processing Soc Workshop Madison WI August 1999 ISSN 1053-587X
"Blind identification and separation of convolutively mixed independent sources" WANG & ZHENYA Pages 997-1002, Vol 33 IEEE Trans on Aerospace & Electronic Systems, 1997 ISSN 0018-9251

(58) Field of Search
UK CL (Edition S) G4A ACF ACL
INT CL[7] G06F 17/15
ONLINE: WPI, EPODOC, JAPIO, INSPEC

(54) Abstract Title
Digital signal processing method and system

(57) A method and system for blind source separation which uses a measure of temporal predictability to identify a source signal from a mixture of such source signals. The temporal predictability of any signal mixture is defined to be less than that of its component source signal. The temporal predictability is used to recover source signals from a set of linear mixtures of source signals by finding an unmixing matrix that maximises a measure of temporal predictability. This measure of temporal predictability can also be used for blind deconvolution, by finding a deconvolutionfilter which minimises the degree of predictability of deconvolved signals. Unlike conventional methods and systems for source separation and blind deconvolution, such as independent component analysis, the embodiments of the present invention produce a closed form solution and require minimal assumptions regarding the probability density function of source signals.

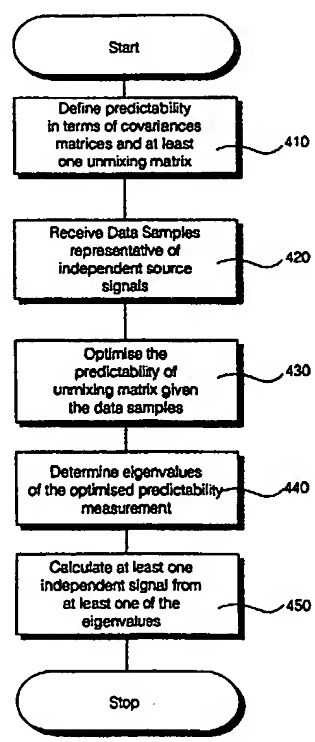


Fig. 4

GB 2 368 936 A

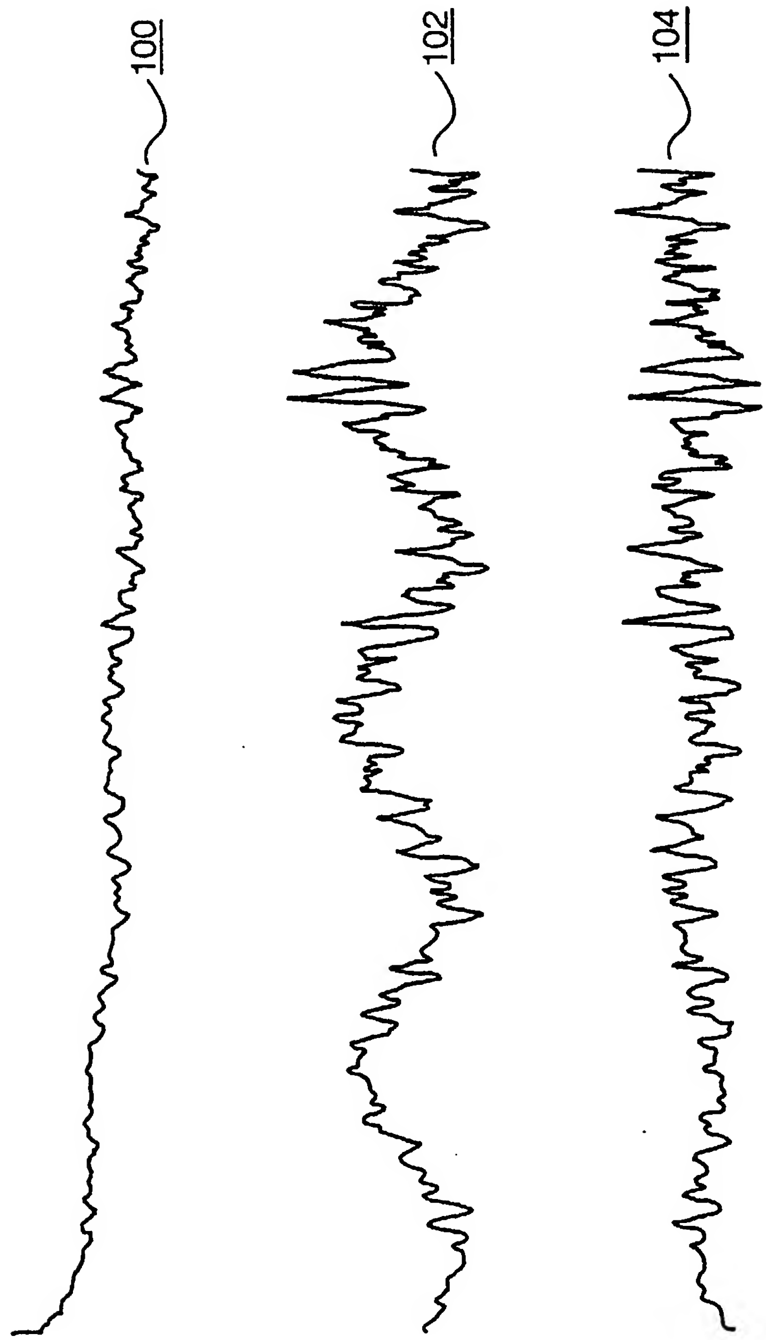
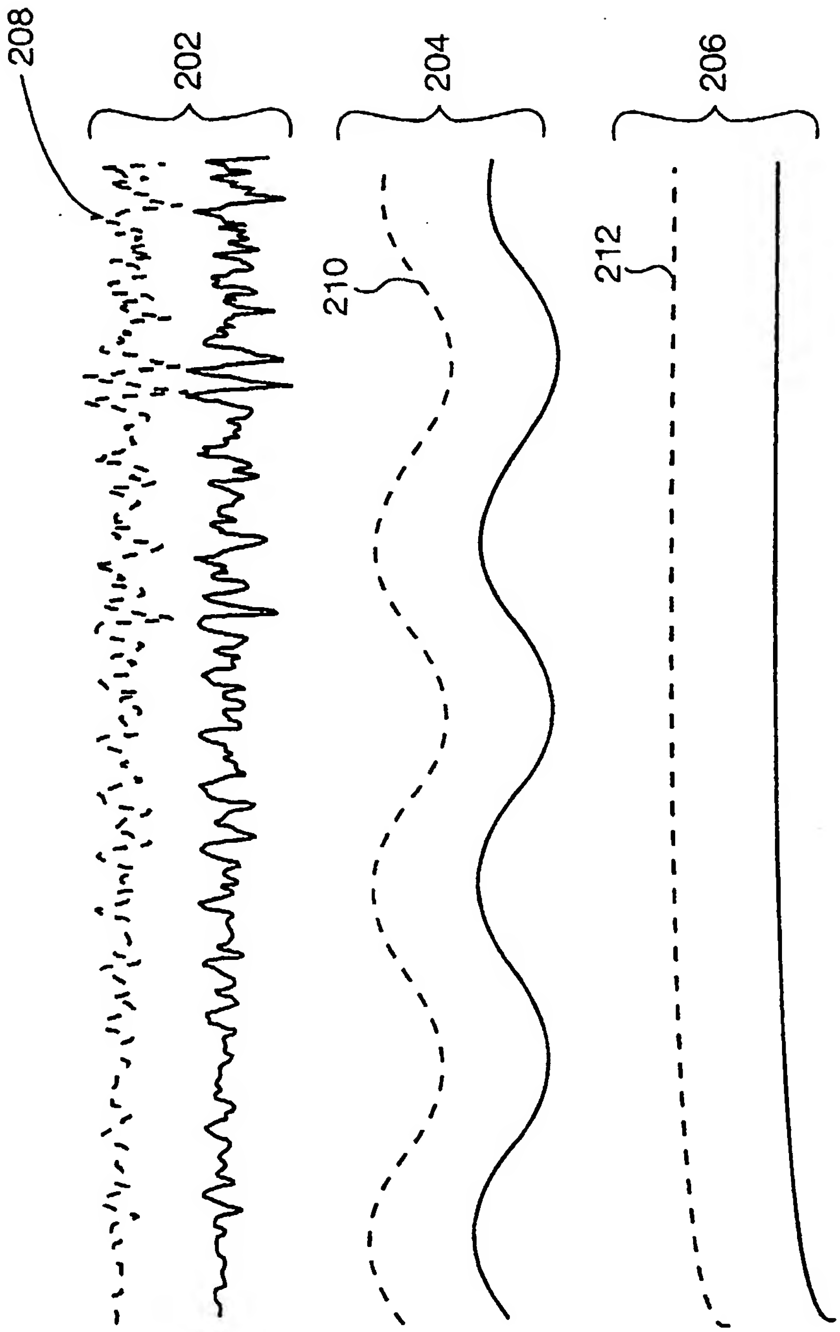


Fig. 1



200

Fig. 2

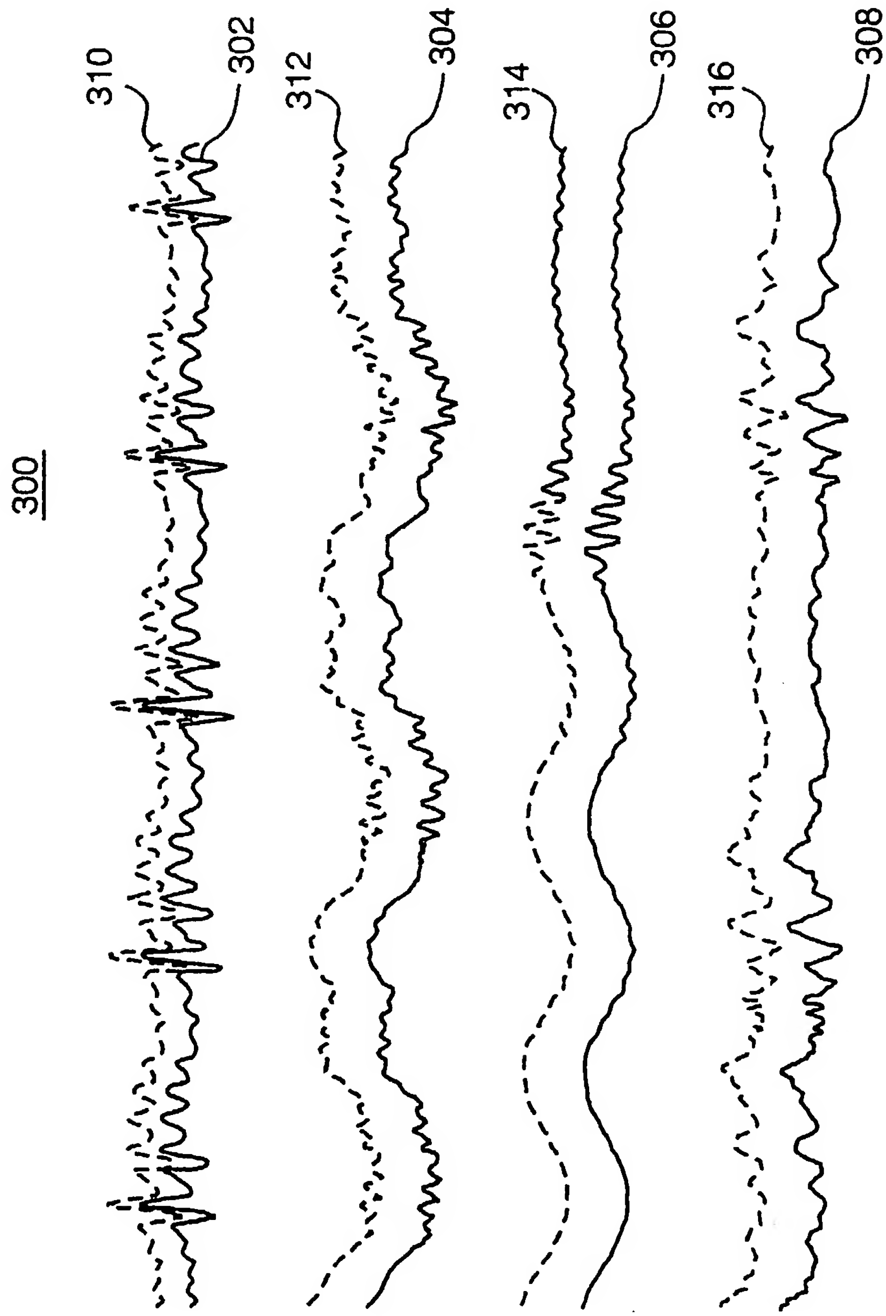


Fig. 3

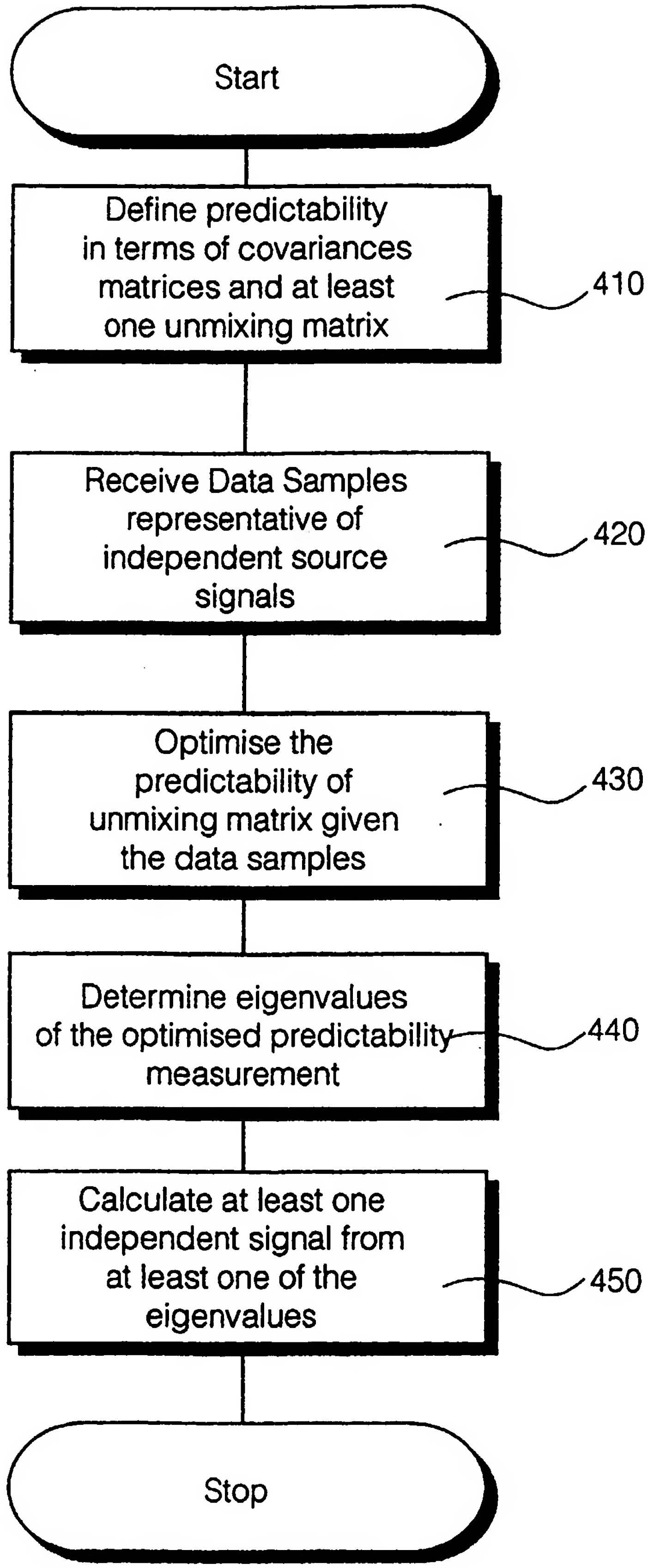

Start
Define predictability in terms of covariances matrices and at least one unmixing matrix
410
Receive Data Samples representative of independent source signals
420
Optimise the predictability of unmixing matrix given the data samples
430
Determine eigenvalues of the optimised predictability measurement
440
Calculate at least one independent signal from at least one of the eigenvalues
450
Stop

Fig. 4

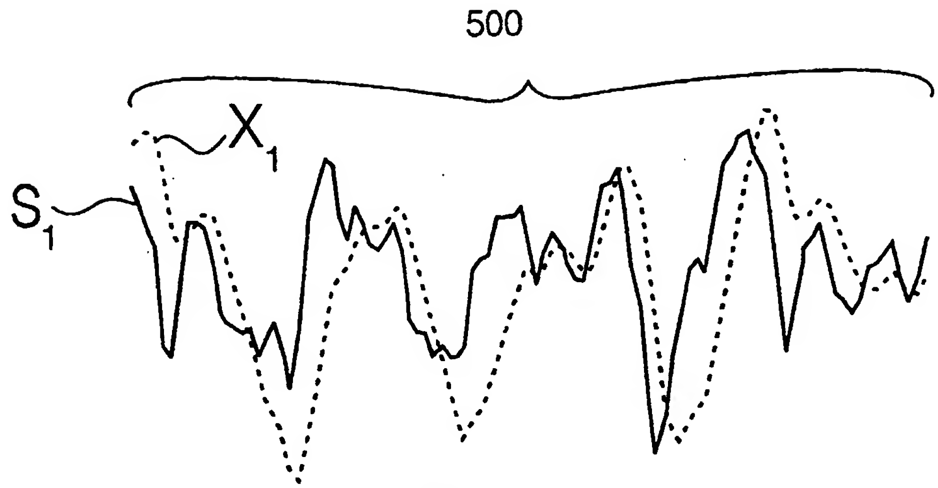


Fig. 5

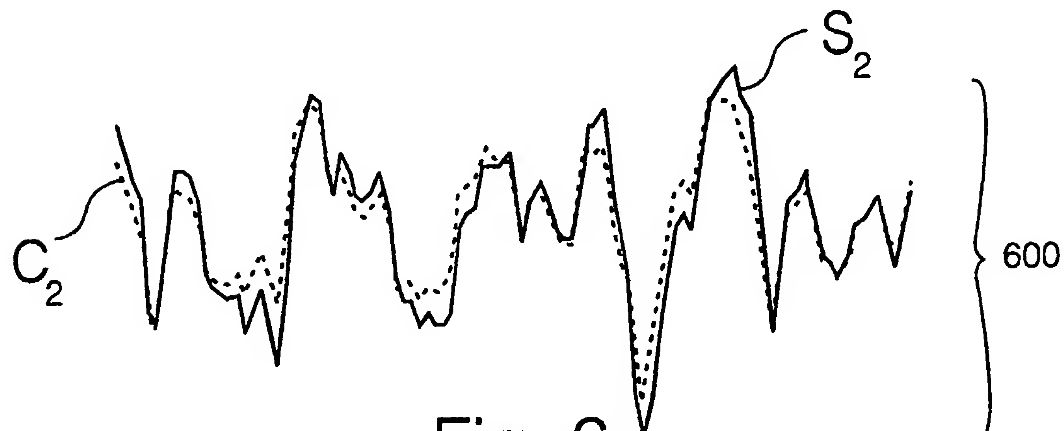


Fig. 6

DIGITAL SIGNAL PROCESSING METHOD AND SYSTEM

The present invention relates to digital signal processing and, more particularly, to blind source signal separation.

In many signal processing applications, the sample signals provided by suitable detectors are mixtures of many unknown signals from unknown sources. The "*separation of sources*" problem is to extract at least one of the unknown signals from such a mixture. Typically, the signal sources as well as the characteristics of the combined mixture signal are unknown. Such signal processing, without knowledge of the signal sources, other than the general statistical assumption of source independence, is known within the art as the "*blind source separation problem*".

The blind source separation problem can arise and have utility in variety of context such as, for example the separation of radio or radar signals sent by an array of antennas, separation of odours in a mixture by a suitable sense array, the parsing of the environment into separate objects by our biological visual system and the separation of biomagnetic sources by a super conducting quantum interference device array in magnetoencephalography. The blind source separation problem finds particular application in, for example, sonar array signal processing and signal processing within a telecommunications system and in particular within a cellular telecommunication system. An example of a context in which a blind source separation problem may arise is two people speaking simultaneously with

each person being a different distance from two microphones. Each microphone records a linear mixture for the two voices. The separation of those voices from the combined signal without knowing the characteristics of the source of the signals would represent such a blind source separation problem.

Blind signal processing or blind source separation are terms used within the art for the recovery of unknown source signals from a received mixture of composite signal. US 5,706,402 discloses a "*Blind signal processing system employing information maximisation to recover unknown signals through unsupervised minimisation of output redundancy*". A neural network system and unsupervised learning process is disclosed for separating unknown source signals from their received mixtures by solving a blind source separation problem. The unsupervised learning procedure solves the general blind signal processing problem by maximising joint output entropy through gradient ascent. The neural network system can separate a multiplicity of unknown source signals from measured mixture signals where the mixture characteristics and the original source signals are both unknowns. However, the system disclosed in US 5,706,402 does not provide an analytical solution and accordingly, is a relatively slow process.

It is an object of the present invention to mitigate at least some of the problems of the prior art.

Accordingly, a first aspect of the present invention provides a digital signal processing method (and system) for deriving a signal from a mixture signal comprising a

combination of a plurality of independent digital signals, the method comprising the steps of

storing a first data set ($\boldsymbol{x}$) representing sampled digital data of the mixture signal;

defining a measure of the predictability of the first data set, the measure being a function of the first data set ($\boldsymbol{x}$) and a first operand (W_i);

processing the first data set representing digital data of a sampled mixture signal comprising a combination of a plurality of a source signals to produce an output signal (y_i) which results from a critical point of the variation of the measure of predictability with the first operand such that the measure of predictability of at least one of the source signals is equal to or greater than the measure of predictability of the sampled mixture signal

Preferably, an embodiment provides a method wherein the definition of the measure of predictability is

$$F = \log\frac{V}{U} = \log\frac{\sum_{\tau=1}^{n}\left(\bar{y}_\tau - y_\tau\right)^2}{\sum_{\tau=1}^{n}\left(\tilde{y}_\tau - y_\tau\right)^2}$$

where y_τ is the value of the signal y at time τ, $\tilde{y}_\tau$ is a short-term moving average of values of y; $\bar{y}_\tau$ is a long-term moving average of values of y; U, the denominator, is a measure of the extent to which y_τ is predicted by the short-term moving average, $\tilde{y}_\tau$, of the past values of y; and V, the numerator, is a measure of the variability of y in terms of the extent to which y_τ is predicted by the long-term moving average, $\bar{y}_\tau$, of the past values of y.

An embodiment provides a method wherein the short-term moving average of the values of y is given by $\tilde{y}_\tau = \lambda_S \tilde{y}_{(\tau-1)} + (1-\lambda_S) y_{(\tau-1)}$ where $0 \le \lambda_S \le 1$, where λ_S is arranged to have a first predetermined half-life, h_S and in which the long-term moving-average of the values of y is given by $\overline{y_\tau} = \lambda_L \overline{y}_{(\tau-1)} + (1-\lambda_L) y_{(\tau-1)}$ where $0 \le \lambda_L \le 1$, where λ_L is arranged to have a second predetermined half-life, h_L. The relationship between a half-life, h, and the parameter λ, is given by $\lambda = 2^{-1/h}$.

Preferably, the second predetermined half-life is longer than the first predetermined half-life and, more preferably, the second predetermined half-life is at least 100 times longer than the first predetermined half-life.

A second aspect of the present invention provides a method (and system) wherein the definition of the measure of predictability is $F = \log \frac{W_i \overline{C} W_i'}{W_i \tilde{C} W_i'}$, where a scalar signal, y_i, is formed from the application of a $1xM$ matrix, W_i, to a vector, $\boldsymbol{x}$, representing potential or actual linear mixtures of the first data set, $\overline{C}$ is an MxM long-term covariance matrix of $\boldsymbol{x}$, and $\tilde{C}$ is a short-term covariance matrix of $\boldsymbol{x}$.

Preferably, an embodiment provides a method in which the long-term covariance, $\overline{C}_{ij}$, between the ith and jth mixtures of $\boldsymbol{x}$ are such that $\tilde{C}_{ij} = \sum_\tau^n (x_{i\tau} - \tilde{x}_{i\tau})(x_{j\tau} - \tilde{x}_{j\tau})$ and $\overline{C}_{ij} = \sum_\tau^n (x_{i\tau} - \overline{x}_{i\tau})(x_{i\tau} - \overline{x}_{j\tau})$. Advantageously, the values $(x_{i\tau} - \overline{x}_{i\tau})$

and $(x_{ir} - \tilde{x}_{ir})$ only needs to be calculated once for a given set of mixture signals.

The technique of gradient ascent can be used to determine the critical points (e.g. saddle points). Suitably, an embodiment provides a method further comprising the step of maximising F with respect to W_i, which gives $\nabla_{W_i} F = \frac{2W_i}{V}\overline{C} - \frac{2W_i}{U}\tilde{C}$.

Preferably, the determination of critical points is an interative process. Accordingly, an embodiment provides a method further comprising the steps of iteratively optimising F with respect in to W_i, using $W_i = W_i + \eta \nabla_{W_i} F$, where η has a predeterminable value. A preferred value of η is 0.001

Having determined the maxima, an embodiment further comprises the step of calculating $y_i = W_i x$ from the value of W_i for which F was at a maxima.

Ultimately, an embodiment may output the recovered signal in a physical form. Accordingly, an embodiment provides a method further comprising the step of outputting a signal based on y_i.

Preferably, an embodiment provides a method further comprising the step of calculating eigenvectors, W_i, for the critical points of F. The step of calculating the eigenvectors, W_i, comprises calculating the eigenvalues, $\lambda = V/U$, where $\nabla_{W_i} F = 0$, which gives $W_i \overline{C} = \frac{V}{U} W_i \tilde{C}$, which, in terms of the eigenvalues, λ, gives $W_i \overline{C} = \lambda W_i \tilde{C}$.

Advantageously, embodiments of the present invention provide a solution to an eigenvalue problem, which if the number of eigenvalues is less then 5, is an analytical solution to the problem of blind source separation by utilising temporal predictability.

The result of voice mixtures described above exemplify three universal properties of linear mixtures of statistically independent source signals which are: (1) temporal predictability (conjecture): the temporal predictability of any signal mixture is less than (or equal to) that of its components or signals, (2) Gaussian probability density function: the central limit theorem ensures that the extent to which the probability density function (pdf) of any mixture approximates a Gaussian distribution is greater than or equal to any of its component source signals, and (3) statistical independence: the degree of statistical independence between any two signal mixtures is less than (or equal to) the degree of independence between any two source signals. Property 2 forms the basis of projection pursuit as disclosed in Freedman, JH. 1987, exploratory projection suit. J amer. Statistical association, **82** (397), 294-266. Properties 1 and 2 are assumptions underlying independent component analysis as disclosed in, for example, Jutten & Herault, 1998, independent component analysis -v- PCA, pages 643-646 of: Proc. Eusipco. and Bell, AJ and Sejnowski, TJ, 1995, and information - maximisation approach to blind separation and blind deconvolution, Neural Computation, **7**, 1129-1159 the entire contents of which are incorporated herein for all purposes. All three properties are generic characteristics of signal mixtures. However, unlike

properties 2 and 3, property 1, that is, temporal predictability forms the only basis of blind source separation for the embodiments of the present invention.

The embodiments of the present invention are based upon the above temporal predictability conjecture which is that the temporal predictability of any mixture is less than (or equal to) that of any of its component signals. Examples will be given below of the use of the temporal predictability conjecture for separation of physic signals, such as, for example, voices and music.

Problem definition and temporal predictability

Before describing an embodiment of the present invention, the theory upon which the embodiments are based will now be described.

Consider a random vector $\mathbf{s} = (s_1|s_2|\ldots|s_K)^t$ of K statistically independent source signals, where the *i*th row in $\mathbf{s}$ is a signal s_i measured at n time points and the superscript t denotes a transpose operator. In the theory presented hereafter it is assumed that the source signals are statistically independent unless otherwise indicated. A random vector $\mathbf{x}=(x_1|x_2|\ldots|x_M)^t$ of $M \geq K$ linear mixtures of signals in s can be formed with an M x K mixing matrix A such that $\mathbf{x} = A\mathbf{s}$. If the rows of A are linearly independent then any source signal s_i can be recovered from $\mathbf{x}$ with a *1xM* matrix, W_i, such that $s_i=W_i\mathbf{x}$. The blind source separation problem to be addressed here consists in finding an unmixing matrix $W = (W_1|W_2|\ldots|W_K)^t$ such that each row vector W_i recovers a different signal

y_i, where y_i is a scaled version and/or sign reversed version of a source signal s_i.

The embodiments of the present invention for recovering source signals are based on the following conjecture: the temporal predictability of a signal mixture x_i is usually less than that of any of the source signals that contribute to x_i. For example, the waveform obtained by adding two sine waves of different frequencies is more "complex" than either of the original sine waves. A measure of temporal predictability, $F(W_i,\mathbf{x})$ is defined and used to estimate the relative predictability of a signal y_i recovered by given matrix W_i, where $y_i=W_i\mathbf{x}$. If source signals are more predictable than any linear mixture, y_i of those signals, then the value of W_i which maximises the predictability of an extracted signal, y_i, should yield a source signal, that is, $y_i = cs_i$, where c is a constant. Information theoretic analysis of F shows that maximising the temporal predictability of a signal amounts to differentially maximising the power of the Fourier components with low (non-zero) frequencies see, for example, Stone J V, 1996a, a Canonical microfunction for learning perceptual invariances, Perception, **25** (2), 207-220 the entire contents of which are incorporated herein for all purposes. In an embodiment, the definition of signal predictability, F, used herein is

$$F = \log\frac{V}{U} = \log\frac{\sum_{\tau=1}^{n}\left(\overline{y_\tau} - y_\tau\right)^2}{\sum_{\tau=1}^{n}\left(\tilde{y}_\tau - y_\tau\right)^2}$$

where y_τ is the value of the signal, y at time τ. The term U measures the extent to which y_τ is predicted by a short-term moving average $\tilde{y}$ of values in y. In contrast, the term V is a measure of the overall variability in y, as measured by the extent to which y_τ is predicted by a long-term moving average y_τ with straight line above the y of values in y. The predicted values $\bar{y}_\tau$ and $\tilde{y}_\tau$ are both exponentially weighted sums of signals values measured up to time τ, such that recent values have a larger weighting then those in the distant past. Therefore, it follows that $\tilde{y}_\tau = \lambda_S \tilde{y}_{(\tau-1)} + (1-\lambda_S) y_{(\tau-1)}$

where $0 \le \lambda_S \le 1$

$$\overline{y_\tau} = \lambda_L \bar{y}_{(\tau-1)} + (1-\lambda_L) y_{(\tau-1)} \qquad \text{where } 0 \le \lambda_L \le 1$$

The half-life h_L of λ_L is much longer (typically of the order of 100 times longer) than the corresponding half-life h_S of λ_S. It should be noted that maximising only V would result in a high variance signal with no constraints on the temporal structure of the signal. In contrast, minimising only U would result in a DC signal. It will be appreciated that in both cases, trivial solutions would be obtained for W_i because V can be maximised by setting the norm of W_i to be large, and U can be minimised by setting $|W_i| = 0$. In contrast, the ratio V/U can be maximised only if two constraints are both satisfied which are (1) y has a non-zero range, that is, a high variance, and (2) the values in y change relatively slowly over time. It should also be noted that the value of F is independent of the norm of W_i, so that only changes in the direction of W_i effect the value of F.

Extracting Signals by Maximising Signal Predictability

Consider a scalar signal , y_i, formed from the application of a *1xM* matrix W_i to a random vector, **x**. Given that $y_i = W_i\mathbf{x}$, equation (1) can be written as:

$$F = \log\frac{W_i\overline{C}W_i'}{W_i\tilde{C}W_i'} \qquad (3)$$

where $\overline{C}$ is an *MxM* long-term covariance matrix of signal mixtures and $\overline{C}$ is a corresponding matrix of short-term covariances. A long-term covariance $\overline{C}_{ij}$ and a short-term covariance $\overline{C}_{ij}$ between an ith and jth mixtures are defined as:

$$\tilde{C}_{ij} = \sum_{\tau}^{n}\left(x_{i\tau} - \tilde{x}_{i\tau}\right)\left(x_{j\tau} - \tilde{x}_{j\tau}\right) \qquad (4)$$

$$\overline{C}_{ij} = \sum_{\tau}^{n}\left(x_{i\tau} - \overline{x}_{i\tau}\right)\left(x_{i\tau} - \overline{x}_{j\tau}\right)$$

It should be noted that the embodiments advantageously only need to compute $\tilde{C}$ and $\overline{C}$ once for any given set of signal mixtures and that the terms $(x_{i\tau} - \overline{x}_{i\tau})$ and $(x_{i\tau} - \tilde{x}_{i\tau})$, can be precomputed using fast convolution operations such as are disclosed in, for example, Eglen, S, Bray, A, and Stone, J V, 1997, Unsupervised discovery of invariances, Network, **8**, 441-442, the entire contents of which are incorporated herein for all purposes.

A gradient ascent on F with respect to W_i can be used to maximise F, thereby maximising the predictability of y_i. The derivative of F with respect to W_i is:

$$\nabla w_i F = \frac{2W_i}{V}\overline{C} - \frac{2W_i}{U}\widetilde{C} \qquad (5)$$

An optimisation procedure is applied which consists of iteratively updating W_i until a maximum F is located such that $W_i = W_i + \eta \nabla_{W_i} F$, where η is a small constant that is, preferably, 0.01. It will be appreciated that the function F is a ratio of quadratic forms. Therefore, F has exactly one global maximum and exactly one global minimum, with all other critical points being saddle points. Therefore, the gradient ascent is guaranteed to find the global maximum in F.

Once a single source signal has been extracted, the repeated application of the above procedure to a single set of mixtures extracts the same (most predictable) source signal. Therefore, it will be appreciated that embodiments are required to ensure that different source signals are extracted. Preferably, the embodiments are arranged to achieve this aim using an unmixing matrix W obtained as the solutions to a generalised eigenproblem defined as follows:

the gradient of F defined in equation 5 is known to be zero at an eigenvalue. Therefore, at an eigenvalue solution

$$W_i\overline{C} = \frac{V}{U} W_i\widetilde{C} \qquad (6)$$

$$W_i\bar{C} = \lambda W_i\tilde{C} \tag{7}$$

It can be appreciated that equations 6 and 7 have the form of a generalised eigenproblem, where the eigenvalues, $\lambda = V/U$, provide the solution. Critical points in F correspond to values of W_i that satisfy equation 7 where W_i's are the corresponding eigenvectors. The first such eigenvector defines a maximum in F and each of the remaining eigenvectors define saddle points in F. It should be noted that these eigenvectors are orthogonal in the metrics defined by $\bar{C}$ and $\tilde{C}$ which implies that different W_i's are linearly independent. Borga M 1998, Learning multidimensional signal processing, Linkoping University, which is incorporated herein for all purposes, discloses a review of generalised eigenproblems.

It will be appreciated by those skilled in the art that such problems typically have scaling characteristics of $O(N^3)$, where N is the number of signal mixtures, that is, the speed of the determination of solutions varies with the cube of the number of independent sources. Once the eigenvectors have been obtained, all K signals can be recovered using $\mathbf{y} = W\,\mathbf{x}$, where each row of $\mathbf{y}$ corresponds to exactly one extracted signal y_i which, as indicated above, represents a scaled version of a corresponding independent source signal s_i.

Preferably, an embodiment provides, in circumstances where the number of mixtures, M, is greater than the number of sources, K, for the reduction of M using principal component analysis as is well known within the

art. Principal component analysis is used to reduce the dimensionality of the signal mixtures by discarding eigenvectors of **x** which have eigenvalues close to zero.

Embodiments of the present invention will now be described, by way of example only, with reference to the accompanying drawings in which:

figure 1 depicts three signal mixtures to be processed by an embodiment of the present invention;

figure 2 shows three signals having different probability density functions of three source signals that were used to synthesise the three signal mixtures shown in figure 1;

figure 3 illustrates signals corresponding to two male and two female voices where the original source mixed signal is shown by continuous lines and corresponding recovered signals shown as dotted lines;

figure 4 shows a flow chart for implementing a method according to an embodiment of the present invention;

figure 5 there is shown two signals comprising an original signal, s_1, denoted by the solid line and a convolved signal, x_1, denoted by the broken line, for 100 time steps processed according to the present invention; and

figure 6 depicts the correlation between an original source signal, s_2, and a deconvolved signal, c_2.

Referring to figure 1 there is shown three mixed signals 100, 102 and 104 each comprising 3000 samples (although only the first 1000 samples of each mixture are shown in figure 1). The three signals were derived from the signals 200 shown in Figure 2. The three source

signals, $\mathbf{s} = \{s_1, |s_2, |s_3\}^t$, shown in figure 2 represent a super-Gaussian signal (such as the sound of train-whistle) 202, a sub-Gaussian signal, such as, for example, a sign wave, 204 and a sorted Gaussian noise signal 206. Each of the signals 202 to 206 were used to synthesise the mixture signals 100, 102 and 104 shown in figure 1 using a random matrix A applied to the set of three signal mixtures given by a $\mathbf{x} = A\mathbf{s}$.

During the processing of the mixture signals 100 to 104, the K source signals, where K = 3 in the present embodiment, were used to generate M = K mixture signals. Again M = 3, using a *KxK* mixing matrix (A) and the M mixture signals were used as input signals for an embodiment. Preferably, each mixture signal was normalised to have zero-mean and unit variance. Each mixing matrix was obtaining using, for example, a random number generator function. The short-term and long-term half-lives defined in equation 2 above were set to $h_S = 1$ and $h_L = 9000$, respectively. Referring again to figure 2, the signals shown by dotted lines, that is, signals 208, 210 and 212 represent amplitude or dc shifted recovered signals which, as can be seen from figure 2, correlate strongly with the original source signals notwithstanding having been mixed by a random matrix A. Table 1 below shows the correlation between the source signals and the signals recovered from the mixtures of the source signals with different probability density functions.

Table 1

	Source 1	Source 2	Source 3
Recovered 1	0.0003	0.0011	**1.0000**

Recovered 2	**1.0000**	0.0000	0.0000
Recovered 3	0.0417	**0.9991**	0.0016

It can be appreciated that the three recovered signals each had a correlation of r>0.99 with only one of the original source signals whereas the other correlations were close to 0. Therefore, embodiments of the present invention can be used to separate mixtures of source signals that have different probability density functions.

An embodiment of the present invention was applied to the separation of sounds where 50,000 data points were sampled at a rate of 44100 Hz, using a microphone to record different voices from a VHF radio using a computer. Two sets of 8 sounds were recorded. The two sets represented voice and music. The voices were a mix of male and female voices. The music was classical, with and without accompanying singing. An embodiment was tested on mixtures of normal speech. Referring to figure 3 there is shown a number of signals 300 which show as complete lines the original source signals 302 to 308 and amplitude or dc shifted versions of the recovered signal 314 to 316 as dotted lines. Table 2 below shows the correlation between each of the four source signal (voices) and every signal recovered from the method. It can be appreciated that each source signal has a high correlation with only one recovered signal.

Table 2

	Source 1	Source 2	Source 3	Source 4

Recovered 1	0.0973	**0.9938**	0.0281	0.0488
Recovered 2	**0.9963**	0.0809	0.0124	0.0191
Recovered 3	0.0015	0.0419	**0.9946**	0.0953
Recovered 4	0.0295	0.0756	0.1014	**0.9916**

Table 3 below shows the correlation between each of eight source signals (voices) and every signal recovered using an embodiment. Again, it can be appreciated that each source signal has a high correlation with only one recovered signal. It can be appreciated that with correlations of this magnitude, it is not possible to discern audibly the difference between the original and the recovered speech signals. It can be seen from Table 3 that a correlation of 0.956 was found between source signal 8 and the recovered signal 3. This, in this example, represents a worse case performance of the method for all data processed according to the above-described embodiment of the present invention.

Table 3

	Source 1	Source 2	Source 3	Source 4	Source 5	Source 6	Source 7	Source 8
Recovered 1	0.0014	0.0076	0.0037	0.0026	0.0278	0.0456	0.9881	0.1430
Recovered 2	0.9936	0.0128	0.0029	0.0013	0.0011	0.0172	0.0155	0.1093
Recovered 3	0.1790	0.0008	0.1612	0.0110	0.0366	0.1020	0.1338	0.9560
Recovered 4	0.0150	0.0118	0.0065	0.9990	0.0240	0.0317	0.0044	0.0044
Recovered 5	0.0035	0.0204	0.9934	0.0002	0.0207	0.0082	0.0074	0.1088
Recovered 6	0.0103	0.0034	0.0264	0.0176	0.0208	0.9919	0.0442	0.112
Recovered 7	0.0267	0.9994	0.0116	0.0015	0.0003	0.0102	0.0088	0.0021
Recovered 8	0.0151	0.0027	0.0269	0.0215	0.9979	0.0197	0.0208	0.0431

The embodiment was also tested using mixtures of eight segments of music. The correlations between the

source signals of the eight segments of music and the recovered signals for the eight segments of music are shown below in Figure 3. It can be appreciated, again, that the correlations are approximately $R = 0.99$ and it is not possible to discern audibly the difference between the original and recovered music signals.

It has been found that the embodiments are largely insensitive to the values used for the short-term and long-term half-lives defined in equation 2 above providing the latter is much greater than the former.

It will be appreciated from the above that the method used in the embodiments is based on the assumption that different source signals are associated (via W_i) with distinct critical points in F. However, if any two source signals have the same degree of predictability, F, then two eigenvectors W_i and W_j will also have equal eigenvalues (and are associated with the same critical points in F). Therefore, any vector W_k which lies in the plane defined by W_i and W_j also maximises F, but W_k cannot, in general, be used to extract a source signal. This can be demonstrated by creating two mixtures $\mathbf{x} = A\,\mathbf{s}$ from two signals s_1 and s_2, where s_1 is a time reversed version of s_2. Even though s_1 and s_2 have different time courses, they share exactly the same degree of predictability F and cannot be extracted from the mixture $\mathbf{x}$ using this method. In practice, however, signals from different sources typically can be separated because each source signal has a unique degree of predictability. Furthermore, every set of signals in which each signal is from a physically distinct source has been successfully separated using embodiments of the present invention.

Although the above embodiments have been described with reference to processing single dimension data, the present invention is not limited thereto. The method can equally well be utilised to realise embodiments which process spatially distributed or N-dimensional data, that is, the method can be applied to realised embodiments that operate in the spatial domain. The definition of predictability can be generalised by replacing the exponentially awaited means $\tilde{y}_\tau$ and $\bar{y}_\tau$ with general functions f and g respectively, where the values of f and g at time τ are non-linear functions of y computed up to time τ. These functions may implement more accurate predictions of y_τ for specific signal types.

It will be appreciated that if all three properties listed above apply to any statistically independent source signals and their mixtures, an embodiment can be realised which relies on constraints from all of these properties to deal with a wide-range of signal types. It is widely acknowledged that ICA forces statistical independence on recovered signals, even in the underlying source signals are not independent. Similarly, the embodiments of the present invention may impose temporal predictability on recovered signals even where none exists in the underlying source signal. Therefore, an embodiment which incorporates constraints from all three properties should be relatively insensitive to violations of the assumptions upon which the embodiments are based. A framework for incorporating experimentally relevant constraints based on physically realistic properties has been formulated in the form of weak models, see, for example, Porrill, J, Stone, J V, Berwick, J, Mayhew, J,

and Coffey, P, Analysis of Optical Imaging Data using Weak Models and ICA, in: Girolami M, (ED) Advances in Independent Component Analysis, Springer - Verlag, which is incorporated herein by reference for all purposes.

It will be appreciated that an application of the present invention may be in the analysis of medical images and EEG data. Alternatively, the present invention could be applied within a hearing aid context in which a selected signal, voice signal, having predeterminable characteristics can be separated from a mixture of voice signals.

Referring to figure 4, in broad terms, there is shown a flow chart 400 for implementing a method according to an embodiment of the present invention. A measure of predictability, F, is defined at step 410 in terms of covariance matrices and at least one unmixing matrix, W_i. Preferably, the definition relates to the definition of F above. Data samples representing a sampled mixture signal which, in turn, is a combination of a plurality of independent source signals is received at step 420 and, optionally, stored in a memory of a digital signal processor. The expression representing the predictability measurement is optimised to determine local critical points in terms of the unmixing matrix, W_i, and the given data samples at step 430. The eigenvalues and, in turn, eigenvectors, of the optimised predictability measurement are determined at step 440 using, preferably, an optimisation procedure. Step 450 calculates using the above equations an independent signal from at least one of the eigenvalues.

Although the embodiments have been described above in relation to a method of processing digital data samples derived from or representing physical signal measurements, the present invention is not limited thereto. Embodiments of the present invention encompass a digital signal processing system or processor that is arranged to implement the above-described methods as well as computer program elements and computer program products for implementing such methods or comprising software for implementing such methods.

Although the above embodiments have been described with reference to recovering source signals are based on the following conjecture: the temporal predictability of a signal mixture x_i is usually less than that of any of the source signals that contribute to x_i, the present invention is not limited thereto. Embodiments can be realised in which the continuity predictability of a signal is used as the measure of predictability such that the continuity predictability of a signal mixture is less than that of any of the source signals that contribute to the mixture signal. Accordingly, the present invention is not limited to temporal continuity and encompasses spatial continuity and n-dimensional continuity.

Deconvolution

The above embodiments have been described with reference to the separation of blind sources. However, the present invention is not limited thereto. Embodiments can be realised in which deconvolution of signals can be performed.

As will be appreciated by one skilled in the art, a signal can be altered by the effects of the environment in which the signal propagates. For example, a room typically causes speech signals to be heard with attenuated high frequencies and with multiple echoes. The room effectively acts as a filter on the signal. Such effects can be reversed using appropriate deconvolution filters. Suitably, embodiments of the present invention are arranged to implement deconvolution filters using the above described temporal predictability.

Firstly, the mathematics underlying the present invention will now be described. Consider a filtered version, x, of a signal, s, such that:

$$x = \alpha * s \qquad (1)$$

where a is a filter $\alpha = [\alpha_1, \alpha_2, \ldots]$. The convolution operator * is defined by :

$$x_t = \alpha_1 s_{t-1} + \alpha_3 s_{t-2} + \ldots \qquad (2)$$

where the subscript t denotes time. If a deconvolution filter, $\beta = [\beta_1, \beta_2, \ldots]$, exists for a then $y = s$ can be recovered from x using

$$y = \beta * x \qquad (3)$$

However, in practice, the correct deconvolution filter cannot be obtained exactly, so that $y \approx s$. Therefore, the embodiments of the present invention use the deconvolution strategies discussed below.

A strategy for Deconvolution

The strategy used to estimate the deconvolution filter B will be described for a simple example. Consider an IIR filter, a, which is a low-pass filter:

$$\alpha_n = e^{-n/T} \qquad (4)$$

where T is the half-life of the exponential. The effects of such IIR filters can be reversed using a simple FIR deconvolution filter:

$$\beta = [1, -e^{-1/T}] \qquad (5)$$

A low-pass filter, such as α, effectively smooths *s* to yield x = a*s, that is the relatively high frequency components are removed from s. This smoothing operation inevitably makes x more predictable than s. In order to emphasise this, consider the limiting case as $T \to \infty$. In the limit, x has a constant value, and is therefore very predictable. It follows that a deconvolution filter β which makes x less predictable may be able to approximately recover s.

Therefore, β is defined as that filter which at least reduces, and preferably minimises, a measure of predictability of $y = a * x$. The differences between the mathematics for deconvolution according to embodiments of the present invention and the above-described blind-source separation is as follows. Given a set of signal mixtures, a linear combination of those mixtures usually

exists such that the original signals can be recovered. It was shown above that such a linear combination could be obtained by maximising the predictability of extracted signals. In contrast, it is shown below that minimising predictability is sufficient to deconvolve a signal.

The convolution operations defined above need to be reformulated in terms of mathematically equivalent vector-matrix operations. Consider a deconvolution filter $\beta = [\beta_1, \beta_2]$, such that $y = \beta * x$. This can be re-written in vector-matrix notation $y = \beta\mathbf{x}$ by defining a vector variable x as:

$$\mathbf{x} = \{x \mid xz^{-1}\}^t \tag{6}$$

where each row of $\mathbf{x}$ is a time-shifted version of x, and the superscript t is the transpose operator. The shift operator z^{-k} is defined by

$$x_{t-k} = x_t z^{-k} \tag{7}$$

The convolution $y=\beta*x$ and the vector matrix multiplication $y=\beta\mathbf{x}$ are exactly equivalent.

<u>Measuring signal predictability</u>

The definition of signal predictability F used here is:

$$F(\beta, x) = \log\frac{V(\beta, x)}{U(\beta, x)} = \log\frac{V}{U} = \log\frac{\sum_t^n (y_t - y_t)^2}{\sum_{t=1}^n (y_t - y_t)^2} \tag{8}$$

where $y_\tau = \beta \mathbf{x}_\tau$ is the value of the signal y at time τ, and $\mathbf{x}_\tau$ is a vector of K signal mixture values of time τ. The term U_i reflects the extent to which y_τ is predicted by a short-term "moving average" $\tilde{y}_\tau$ of values in y. In contrast, the term V_i is a measure of the overall variability in y, as measured by the extent to which y_τ is predicted by a long-term "moving average" $\bar{y}_\tau$ of values in y. The predicted values $\tilde{y}_\tau$ and $\bar{y}_\tau$ of y_τ are both exponentially weighted sums of signal values measured up to time $(\tau-1,)$, such that recent values have a larger weighting than those in the distant past:

$$\tilde{y}_\tau = \lambda_S \tilde{y}_{(\tau-1)} + (1-\lambda_S) y_{(\tau-1)} : 0 \le \lambda_S \le 1$$
$$\bar{y}_\tau = \lambda_L \bar{y}_{(\tau-1)} + (1-\lambda_L) y_{(\tau-1)} : 0 \le \lambda_L \le 1 \qquad (9)$$

The half-life h_L of λ_L is much longer (typically 100 times longer) than the corresponding half-life h_S of λ_S. The relationship between a half-life h and the parameter λ is defined as $\lambda = 2^{-1/h}$.

It should be noted that minimising only Vi would result in a low variance signal with no constraints on its temporal structure. In contrast, maximising only U would result in a signal with arbitrarily high amplitude. In both cases, trivial solutions would be obtained for β because V_i can be minimised by setting the norm of β to be zero and U can be maximised by setting the norm of β to be large. In contrast, the ratio V_i/U_i can be minimised only if two constraints are both satisfied: 1) y has a small non-zero range (i.e. low variance), and 2) the values in y change "quickly" over time relative to h_s. It

should also be noted that the value of F is independent of the norm of β, so that only change in the *direction* of β affect the value of F. Advantageously, conventional iterative gradient descent on F shows that the length of β varies only a little throughout the optimisation process. This is in contrast to the embodiment of the present invention.

Deconvolution Using Temporal Predictability

Given that y=β**x**, equation (8) can be re-written as:

$$F = \log \frac{\beta \overline{C} \beta^t}{\beta \widetilde{C} \beta^t}, \qquad (10)$$

where $\overline{C}$ is an M x M matrix of long-term convariances between signal mixtures, and $\widetilde{C}$ is a corresponding matrix of short-term convariances. The long-term covariance $\overline{C}_{ij}$ and the short-term covariance $\widetilde{C}_{ij}$ between the *i*th and *j*th mixtures are defined as:

$$\widetilde{C}_{ij} = \sum_{\tau}^{n} (x_{i\tau} - \widetilde{x}_{i\tau})(x_{j\tau} - \widetilde{x}_{j\tau})$$

$$\overline{C}_{ij} = \sum_{\tau}^{n} (x_{i\tau} - \overline{x}_{i\tau})(x_{j\tau} - \overline{x}_{j\tau}) \qquad (11)$$

It should be noted that that $\widetilde{C}$ and $\overline{C}$ need only be computed once for a given set of signal mixtures, and that the terms $(x_{i\tau} - \overline{x}_{i\tau})$ and $(x_{i\tau} - \widetilde{x}_{i\tau})$, can be pre-computed using fast convolution operations.

Referring to figure 5 there is shown two signals comprising an original signal, s_1, denoted by the solid

line and a convolved signal, x_1, denoted by the broken line, for 100 time steps.

Figure 6 depicts a further original signal, s_2, denoted by the solid line, and a deconvolved, c_2, signal $y=\beta\mathbf{x}$, denoted by a broken line.

Gradient descent on F with respect to β could be used to minimise F, thereby minimising the predictability of y. The derivative of F with respect to β is:

$$\nabla_\beta F = \frac{2\beta}{V}\overline{C} - \frac{2\beta}{U}\widetilde{C} \qquad (12)$$

In an embodiment, one optimising procedure would consist of iteratively updating β until a minimum of F is located: $\beta = \beta - \eta\nabla_\beta F$, where η is a small constant (typically, η = 0.001).

It should be noted that the function F is a ratio of quadratic forms. Therefore, F has exactly one global maximum and exactly one global minimum, with all other critical points being saddle points. This implies that gradient descent is guaranteed to find the global minimum in F.

Deconvolution as an Eigenproblem

The gradient of F is zero at a solution where, from Equation (12):

$$\beta\overline{C} = \frac{V}{U}\beta\widetilde{C} \qquad (13)$$

Critical points of F correspond to values of β that satisfy Equation (13), which has the form of a generalised eigenproblem (Borga 1998). Solutions for β can therefore be obtained as eigenvectors of the matrix $(\tilde{C}^{-1}\overline{C})$, which has corresponding eigenvalues $\gamma = V/U$. As noted above, the first and last such eigenvector define a maximum and a minimum in *F*, respectively, and each of the remaining eigenvectors define saddle points.

It should be noted that eigenproblems have scaling characteristics of $O(N^3)$, where N is the number of signal mixtures. The vector β can be obtained using a generalised eigenvalue routine.

Given an *n*th order FIR filter, *W* is an *n* x *n* matrix. The deconvolved signal can then be recovered by using the smallest eigenvector β^K in $W : y = \beta^K x$ .W : $y=\beta^K\mathbf{x}$.

Deconvolution Results

The embodiments of the present invention were demonstrated using 5000 samples of a signal, s_1, which was the sound of a choir singing, sampled at 8192Hz. The signal, s_1, was convolved with a low-pass filter with n = 32 coefficients $\alpha = e^{-n/T}$, where T = 8 samples. The resultant convolved signal x_1=a*s was used as input to for the embodiments.

Figure 5 displays a short segments 500 of s_1 and x_1, and the correlation between s_1 and x_1 is r = 0.430. After applying the embodiment described above, the correlation between the deconvolved segments 600, that is, signal

$y=\beta^{*}x$ and c_2 is r = 0.980 as can be appreciated from figure 6. It can be appreciated from figure 6 it can be seen that there is good agreement between s and the recovered signal y.

It will be appreciated from the above that the flowchart shown in figure 4 is equally applicable to deconvolution with the exception that deconvolution, in contrast to the maximisation operations of the bss problem, attempts minimise predictability.

Source Separation for Image Mixtures

Although the above embodiments have been described with reference to processing single dimensional signals, embodiments of the present invention are not limited thereto. Embodiments of the present invention can be adapted for a spatial model, although it remains essentially the same. In the spatial model, an embodiment would maximise a function $F=\log(V|U)$ in which the long-term and short-term variances are spatially defined as

$$V=\sum_{i,j}(\bar{y}_{i,j}-y_{i,j})^2 \text{ and } U=\sum_{i,j}(\tilde{y}_{i,j}-y_{i,j})^2 \qquad (14)$$

where i,j are spatial indices spanning, for example, data representing an extracted image y. The quantities $\bar{y}_{i,j}$ and $\tilde{y}_{i,j}$ represent long-range and short-range exponentially weighted spatial means of y, where the weighting is centred at position i,j.

As in the temporal case, V and U can be computed as:

$$F = \log \frac{W\overline{C}W'}{W\widetilde{C}W'}, \tag{15}$$

and solution vectors $W = (W_1, \ldots, W_K)$ can be obtained using the same eigen decomposition as was used in the temporal case.

However, in the case of 2D images, the long-range covariance $\overline{C}_{i,j}$ and short-range covariance $\widetilde{C}_{i,j}$ between the *i*th and *j*th image mixtures are:

$$\begin{aligned} \widetilde{C}_{i,j} &= \sum_{r,c} \sum_{p,q} (\widetilde{x}^i_{r,c} - x^i_{r-p,c-q})(\widetilde{x}^j_{r,c} - x^j_{r-p,c-q}) \\ \overline{C}_{i,j} &= \sum_{r,c} \sum_{p,q} (\overline{x}^i_{r,c} - x^i_{r-p,c-q})(\overline{x}^j_{r,c} - x^j_{r-p,c-q}) \end{aligned} \tag{16}$$

The quantities $\widetilde{C}_{i,j}$ and $\overline{C}_{i,j}$ are calculated by convolving each image mixture with an exponentially decaying masks $\overline{\Phi}$ and $\widetilde{\Phi}$ of different half lives:

$$\overline{x}_{i,j} = \sum_{r,c=-\infty}^{r,c=\infty} \overline{\Phi}_{r,c} x_{i+r,j+c} \quad \text{and} \quad \widetilde{x}_{i,j} = \sum_{r,c=-\infty}^{r,c=\infty} \widetilde{\Phi}_{r,c} x_{i+c,j+y} \tag{17}$$

It should be noted that the entries in the masks are rotationally symmetric and decay exponentially as a function of distance from the mask centre at *r,c*. The rates of decay in $\overline{\Phi}$ and $\widetilde{\Phi}$ are defined by the spatial short-range and long-range half-lives, h_L and h_S, respectively. Recommended values for these are $h_L = 90000$ pixels and $h_S = 1$ pixels.

Therefore, the main difference between the spatial case and the temporal case is the source of contributions to $\tilde{y}$ and $\bar{y}$. Whereas the temporal version sums outputs of a single unit over time, the spatial version sums the output over a spatial array.

The extracted images $y_1,\ldots,y_K$ are recovered from the set of K image mixtures are:

$$y_1 = w_1^1 x_1 + w_2^1 x_2 + \ldots + w_K^1 x_K \quad (18)$$

where $(w_1^i, w_2^i,\ldots,w_K^i)$ are elements of the ith row vector in W.

It will be appreciated that even though the above embodiment has been described with reference to processing 2-D data representing images, the present invention is not limited thereto. The 2-D data could equally well represent data other than image data.

The above embodiments make reference to the term "critical points". It will be appreciated that the meaning of this term includes maxima, minima and saddle points, see, for example, Penguin Dictionary of Mathematics, J Daintith, R Nelson, 1989.

It will be appreciated from the above that the blind source separation problem requires maximising predictability which implies maximisation of F which, in turn, implies $W_i = W_i + \eta \nabla_{W_i} F$. In contrast, deconvolution requires minimisation of predictability which implies minimisation of F which implies $W_i = W_i - \eta \nabla_{W_i} F$, that is,

where F is a measure of predictability, it is maximised for BSS and minimised for deconvolution.

It will be appreciated by those skilled in the art that the above inventions and embodiments can be implemented using, for example, a conventional computer and appropriate digital signal processing software. Alternatively or additionally, a specific hardware platform may be realised for implementing the inventions or embodiments.

The reader's attention is directed to all papers and documents which are filed concurrently with or previous to this specification in connection with this application and which are open to public inspection with this specification, and the contents of all such papers and documents are incorporated herein by reference.

All of the features disclosed in this specification (including any accompanying claims, abstract and drawings), and/or all of the steps of any method or process so disclosed, may be combined in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

Each feature disclosed in this specification (including any accompanying claims, abstract and drawings), may be replaced by alternative features serving the same, equivalent or similar purpose, unless expressly stated otherwise. Thus, unless expressly stated otherwise, each feature disclosed is one example only of a generic series of equivalent or similar features.

The invention is not restricted to the details of any foregoing embodiments. The invention extends to any novel one, or any novel combination, of the features disclosed in this specification (including any accompanying, abstract and drawings), or to any novel one, or any novel combination, of the steps of any method or process so disclosed.

CLAIMS

1. A digital signal processing method for deriving a signal from a mixture signal comprising a combination of a plurality of independent digital signals, the method comprising the steps of

storing or receiving a first data set representing sampled digital data of the mixture signal;

defining a measure of the predictability of the first data set, the measure being a function of the first data set (**x**) and an first operand (W_i);

processing the first data set representing digital data of a sampled mixture signal comprising a combination of a plurality of a source signals to produce an output signal (y_i) which results from a critical point of the variation of the measure of predictability with the first operand such that the measure of predictability of at least one of the source signals is equal to or greater than the measure of predictability of the sampled mixture signal.

2. A method as claimed in claim 1 wherein the definition of the measure of predictability is

$$F = \log\frac{V}{U} = \log\frac{\sum_{\tau=1}^{n}\left(\bar{y}_\tau - y_\tau\right)^2}{\sum_{\tau=1}^{n}\left(\tilde{y}_\tau - y_\tau\right)^2} \qquad (1)$$

where y_τ is the value of the signal y at time τ,

$\tilde{y}_\tau$ is a short-term moving average of the values of y;

$\overline{y}_\tau$ is a long-term moving average of the values of y;

U, the denominator, is a measure of the extent to which y_τ is predicted by the short-term moving average, $\tilde{y}_\tau$, of the values of y; and

V, the numerator, is a measure of the variability of y in terms of the extent to which y_τ is predicted by the long-term moving average, $\overline{y}_\tau$, of the values of y.

3. A method as claimed in claim 2 wherein the short-term moving average of the currently predicted values of y is given by $\tilde{y}_\tau = \lambda_S \tilde{y}_{(\tau-1)} + (1-\lambda_S) y_{(\tau-1)}$ where $0 \le \lambda_S \le 1$, where λ_S is arranged to have a first predetermined half-life, h_S, where $\lambda_s = 2^{-1/h_s}$.

4. A method as claimed in either of claims 2 and 3 wherein the long-term moving-average of the currently predicted values of y is given by $\overline{y_\tau} = \lambda_L \overline{y}_{(\tau-1)} + (1-\lambda_L) y_{(\tau-1)}$ where $0 \le \lambda_L \le 1$, where λ_L is arranged to have a second predetermined half-life, h_L, where $\lambda_L = 2^{-1/h_L}$.

5. A method as claimed in claim 4 wherein the second predetermined half-life is longer than the first predetermined half-life.

6. A method as claimed in claim 5 wherein the second predetermined half-life is at least 100 times longer than the first predetermined half-life.

7. A method as claimed in any preceding claim wherein the definition of the measure of predictability is $F = \log \frac{W_i \overline{C} W_i'}{W_i \widetilde{C} W_i'}$, where a scalar signal, y_i, is formed from the application of a *1xM* matrix, W_i, to a vector variable, $\boldsymbol{x}$, representing potential or actual linear mixtures of the first data set, $\overline{C}$ is an *MxM* long-term covariance matrix of $\boldsymbol{x}$, $\widetilde{C}$ is a short-term covariance matrix of $\boldsymbol{x}$.

8. A method as claimed in claim 7 in which the long-term covariance, $\overline{C}_{ij}$, between the *i*th and *j*th mixtures of $\boldsymbol{x}$ such that $\widetilde{C}_{ij} = \sum_{\tau}^{n} (x_{i\tau} - \widetilde{x}_{i\tau})(x_{j\tau} - \widetilde{x}_{j\tau})$ and $\overline{C}_{ij} = \sum_{\tau}^{n} (x_{i\tau} - \overline{x}_{i\tau})(x_{i\tau} - \overline{x}_{j\tau})$.

9. A method as claimed in either of claims 7 and 8, further comprising the step of locating critical points of F with respect to W_i, which gives $\nabla_{W_i} F = \frac{2W_i}{V} \overline{C} - \frac{2W_i}{U} \widetilde{C}$.

9. A method as claimed in claim 9 further comprising the steps of iteratively optimising F with respect to W_i, using $W_i = W_i + \eta \nabla_{W_i} F$, where η has a predeterminable value.

10. A method as claimed in claim 9 further comprising the steps of iteratively optimising F with respect to W_I, using $W_i = W_i - \eta \nabla_{W_i} F$, where η has a predetermined value.

11. A method as claimed in either of claims 9 and 10 in which the predeterminable value η is 0.001.

12. A method as claimed in any of claims 7 to 11, further comprising the step of calculating $y_i = W_i x$ from the value of W_i for which F was at a critical point.

13. A method as claimed in claim 12, further comprising the step of outputting a signal based on y_i.

14. A method as claimed in any of claims 7 to 13, further comprising the step of calculating eigenvectors, W_i, for the critical point of F.

15. A method as claimed in claim 14 in which the step of calculating the eigenvectors, W_i, comprises calculating the eigenvalues, $\lambda = V/U$, where $\nabla_{W_i} F = 0$, which gives $W_i \overline{C} = \frac{V}{U} W_i \widetilde{C}$, which, in terms of the eigenvalues, λ, gives $W_i \overline{C} = \lambda W_i \widetilde{C}$.

16. A digital signal processing method substantially as described herein with reference to and/or as illustrated by the accompanying drawings.

17. A digital signal processing system for deriving a signal from a mixture signal comprising a combination of a plurality of independent digital signals, the system comprising memory for storing a first data set representing sampled digital data of the mixture signal; means for defining a measure of the predictability of the first data set, the measure being a function of the first data set (**x**) and an first operand (W_i) such that the measure of

predictability of at least one of the source signals is equal to or greater than the measure of predictability of the sampled mixture signal; a processor for processing the first data set representing digital data of a sampled mixture signal comprising a combination of a plurality of a source signals to produce an output signal (y_i) which results from critical points of the variation of the measure of predictability with the first operand.

18. A system as claimed in claim 17 wherein the definition of the measure of predictability is $F = \log\frac{V}{U} = \log\frac{\sum_{\tau=1}^{n}\left(\bar{y}_\tau - y_\tau\right)^2}{\sum_{\tau=1}^{n}\left(\tilde{y}_\tau - y_\tau\right)^2}$ where y_τ is the value of the signal y at time τ, $\tilde{y}_\tau$ is a short-term moving average of the past values of y; $\bar{y}_\tau$ is a long-term moving average of the past values of y; U, the denominator, is a measure of the extent to which y_τ is predicted by the short-term moving average, $\tilde{y}_\tau$, of the past values of y; and V, the numerator, is a measure of the variability of y in terms of the extent to which y_τ is predicted by the long-term moving average, $\bar{y}_\tau$, of the past values of y.

19. A system as claimed in claim 18 wherein the short-term moving average of the values of y is given by $\tilde{y}_\tau = \lambda_S \tilde{y}_{(\tau-1)} + (1-\lambda_S)y_{(\tau-1)}$ where $0 \le \lambda_S \le 1$, where λ_S is arranged to have a first predetermined half-life, h_S, where $\lambda_s = 2^{-1/h_s}$.

20. A system as claimed in either of claims 18 and 19 wherein the long-term moving-average of the values

of y is given by $\overline{y_\tau} = \lambda_L \bar{y}_{(\tau-1)} + (1-\lambda_L) y_{(\tau-1)}$ where $0 \le \lambda_L \le 1$, where λ_L is arranged to have a second predetermined half-life, h_L, $\lambda_L = 2^{-1/h_L}$.

21. A system as claimed in claim 20 wherein the second predetermined half-life is longer than the first predetermined half-life.

22. A system as claimed in claim 21 wherein the second predetermined half-life is at least 100 times longer than the first predetermined half-life.

23. A system as claimed in any of claims 17 to 22 wherein the definition of the measure of predictability is $F = \log \frac{W_i \bar{C} W_i'}{W_i \tilde{C} W_i'}$, where a scalar signal, y_i, is formed from the application of a *1xM* matrix, W_i, to a vector variable, $\boldsymbol{x}$, representing potential or actual linear mixtures of the first data set, $\bar{C}$ is an *MxM* long-term covariance matrix of $\boldsymbol{x}$, $\tilde{C}$ is a short-term covariance matrix of $\boldsymbol{x}$.

24. A system as claimed in claim 23 in which the long-term covariance, $\bar{C}_{ij}$, between the *i*th and *j*th mixtures of $\boldsymbol{x}$ such that $\tilde{C}_{ij} = \sum_{\tau}^{n} (x_{i\tau} - \tilde{x}_{i\tau})(x_{j\tau} - \tilde{x}_{j\tau})$, and $\bar{C}_{ij} = \sum_{\tau}^{n} (x_{i\tau} - \bar{x}_{i\tau})(x_{i\tau} - \bar{x}_{j\tau})$.

25. A system as claimed in either of claims 23 and 24, further comprising means for locating critical points of F with respect to W_i, which gives $\nabla_{W_i} F = \frac{2W_i}{V}\bar{C} - \frac{2W_i}{U}\tilde{C}$.

26. A system as claimed in claim 25 further comprising means for iteratively optimising F with respect in to W_i, using $W_i = W_i + \eta \nabla_{W_i} F$, where η has a predeterminable value.

27. A system as claimed in claim 25 further comprising means for iteratively optimising F with respect to W_i, using $W_i = W_i - \eta \nabla_{W_i} F$, wherein η has a predeterminable value.

28. A system as claimed in either of claims 26 and 27 in which the predeterminable value η is 0.001.

29. A system as claimed in any of claims 17 to 28, further comprising means for calculating $y_i = W_i x$ from the value of W_i for which F was at a maxima.

30. A system as claimed in claim 29, further comprising means for outputting a signal based on y_i.

31. A system as claimed in any of claims 17 to 30, further comprising means for calculating eigenvectors, W_i, for the critical point of F.

32. A system as claimed in claim 31 in which the means for calculating the eigenvectors, W_i, comprises calculating the eigenvalues, $\lambda = V/U$, where $\nabla_{W_i} F = 0$, which gives $W_i \overline{C} = \frac{V}{U} W_i \widetilde{C}$, which, in terms of the eigenvalues, λ, gives $W_i \overline{C} = \lambda W_i \widetilde{C}$.

33. A digital signal processing method substantially as described herein with reference to and/or as illustrated by the accompanying drawings.

34. A computer program element for implementing a system or method as claimed in any preceding claim.

35. A computer program product comprising a storage medium having stored thereon a computer program element as claimed in claim 34.

The Patent Office

INVESTOR IN PEOPLE

Application No:	GB 0103839.7	41	**Examiner:**	Nigel Hanley
Claims searched:	1-35		**Date of search:**	17 October 2001

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.S): G4A (ACL, ACF)

Int Cl (Ed.7): G06F 17/15

Other: ONLINE: WPI, EPODOC, JAPIO, INSPEC

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
A,E	WO 01/17109 A1	SARNOFF - See whole document. Note solution to Blind Source Separation (BSS) problem using Fourier transforms and correlation	
A	US 5959966 A	MOTOROLA - See whole document. Note specific solution of BSS problem in mobile telephony network using probability density functions to identify a mixing matrix.	
A	US 5825671 A	DEVILLE - See whole document. Note use of convolution in identifying a primary signal amongst a plurality of signals and use in controlling an electrical apparatus such as a car radio.	
A	ZARZOSO & NANDI - "Adaptive blind source separation for virtually any source probability density function" - Pages 477-488, IEEE Transactions on Signal Processing Feb 2000, IEEE, USA. ISBN 0-7803-5673-X - *Note iterative method of using a predicted value with an estimate of the mixing matrix of the source signal until a correlation is achieved.*		

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
		E	Patent document published on or after, but with priority date earlier than, the filing date of this application.
&	Member of the same patent family		

Application No: GB 0103839.7 44
Claims searched: 1-35

Examiner: Nigel Hanley
Date of search: 17 October 2001

Category	Identity of document and relevant passage	Relevant to claims
A	MIQUEZ & CASTEDO - "Maximum Likelihood blind source separation in Gaussian Noise"- Pages 343-352 Proceedings of the 1999 IEEE Signal Processing Society Workshop, Madison WI, USA 23-25 Aug 1999. ISSN 1053-587X- *Note use of maximum likelihood method of solving the mixing matrix in the BSS problem.*	
A	WANG & ZHENYA - " Blind identification and separation of convolutively mixed independent sources" - Pages 997-1002 Vol 33, IEEE Transactions on Aerospace and Electronic Systems, July 1997, IEEE, USA - ISSN 0018-9251- *Note use of probabilities in identifying signal from a source signal.*	

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
		E	Patent document published on or after, but with priority date earlier than, the filing date of this application.
&	Member of the same patent family		